Defeasible Reasoning with Quantifiers

Lupita Estefania Gazzo Castañeda (Estefania.Gazzo@psychol.uni-giessen.de)
Markus Knauff (Markus.Knauff@psychol.uni-giessen.de)
Department of Experimental and Cognitive Psychology, Otto-Behaghel-Straße 10F
D-35394 Gießen, Germany

Abstract

Human conditional reasoning is defeasible: people withdraw logically valid conclusions if they are aware of situations (i.e., exceptions) that prevent the consequent of the rule to happen although the antecedent is given. In this paper we investigate defeasible reasoning with quantified rules. In two experiments we rephrased conditionals from the literature (Experiment 1) and rules from penal code (Experiment 2) as either universal or existential rules and embedded them into Modus Ponens and Modus Tollens inference problems. We show that defeasible reasoning also exists for quantified rules. However, the kind of quantifier (universal vs. existential) did not affect inferences. This last finding conflicts with theories highlighting the importance of logic in human reasoning.

Keywords: Quantifiers; defeasible reasoning; exceptions

Introduction

How do humans reason? This question has kept cognitive psychology busy for several decades. In the beginning this question was investigated by analyzing people's capacity to reason according to classical logic (Evans, 2002). People were confronted with conditional inference tasks and asked to make logical inferences. If they were able to draw logically valid conclusions they were considered rational. Responses not corresponding to classical logic were considered "errors" and a sign of irrationality (Evans, 2002; Oaksford & Chater, 2001). For instance, when people were confronted with a conditional rule such as "If a person stands in the sun (*p*), then the person gets a sunburn (*q*)" together with the fact that the antecedent *p* is given (i.e., a person standing in the sun), then participants had to conclude that the consequent *q* follows (i.e., the person getting a sunburn). This is the valid inference of Modus Ponens (MP) from classical logic. The same was the case for the valid inference of Modus Tollens (MT): when participants were confronted with the same conditional but then with the fact that the consequent is not the case ($\neg q$; i.e., the person does not get a sunburn), then they should conclude that the antecedent is also not the case ($\neg p$; i.e., the person is not standing in the sun).

However, over the years researchers have recognized that when participants deny logically valid conclusions it is not necessarily because of "irrationality", but because everyday reasoning often does not follow the rules of classical logic (Evans, 2012; Oaksford & Chater, 2001). In everyday reasoning also the content of the conditional matters and people introduce their content related background knowledge to inference tasks (Cummins, Lubart, Alksnis, & Rist, 1991; De Neys, Schaeken, & d'Ydewalle 2003a; 2003b; Evans, 2002; Johnson-Laird & Byrne, 2002). For instance, in the example above, participants might refuse to conclude that the person gets a sunburn although the person is standing in the sun if they consider the possibility of using sunscreen or already having a good tan. In everyday reasoning people withdraw otherwise valid conclusions if they can think of information which prevents the consequent *q* to occur although the antecedent *p* is given. This consideration of exceptions shows that contrary to classical logic, everyday reasoning is *defeasible* (e.g., Oaksford & Chater, 2001). The more exceptions people consider during reasoning and the more semantically associated these exceptions are to the consequent of the conditional, the more an otherwise valid conclusion is rejected (Cummins et al., 1991; DeNeys et al, 2003a; 2003b; Quinn & Markovits, 1998). In this way, the more exceptions a person can think of, the less he or she perceives the conditional probability of *q* given *p* (Weidenfeld, Oberauer, & Hörnig, 2005; see also Oaksford & Chater, 2001; 2013; Oaksford, Chater, Larkin, 2000).

Still, as far as we know, all research on the consideration of exceptions in defeasible reasoning has been done with conditionals with the classical logical connective *if-then*. But are exceptions also considered when reasoning with *quantifiers* such as in "*All* people standing in the sun get a sunburn" or "*Some* people standing in the sun get a sunburn"? Some directions can be found in Chater and Oaksford (1999); however the question is still particularly interesting because of the distinction between universal (i.e., All As are Bs) and existential (i.e., Some As are Bs) quantifiers, with the latter already suggesting the existence of exceptions. Is it therefore possible that the effect of exceptions in defeasible reasoning is moderated by the quantifier used in the rule?

The aim of this paper is to investigate 1) whether exceptions are considered when reasoning with quantifiers, and 2) to which extent different quantifiers (universal vs. existential) affect the consideration of exceptions. In our reasoning problems we expected people to consider exceptions when reasoning with quantifiers; making reasoning with quantifiers defeasible. However, we hypothesize that the kind of quantifier used in the rule influences the extent to which exceptions are considered. Based on the logical and linguistic implications of universal and existential quantifiers, we suppose that existential quantifiers trigger the consideration of exceptions in such a way that participants show higher reluctance to accept logically valid conclusions when the rule is phrased with an existential compared to a universal quantifier. Showing that reasoning with quantifiers is defeasible, and that defeasibility depends on the kind of quantifier used, would have important implications for (1) theories highlighting the defeasible nature of human reasoning, and (2) theories emphasizing the role of logics in reasoning.

Our hypotheses are also related to the dual source model of Klauer, Beller, and Hütter (2010). Klauer and colleagues argue that the inferences participants draw depend on their background knowledge about the content of the task and its logical form. Whereas the background knowledge component is influenced by the conditional probability of *q* given *p*, the form component depends on the subjective probability of the logical inference presented in the task, but also on the form of the rule and its validity (Klauer et al., 2010; Singmann, Klauer and Over, 2014). For instance, when Klauer et al. (2010) asked participants to estimate how probable it is that *q* follows from *p*, participants gave higher ratings when they were first confronted with the corresponding "if-then" conditional rule than when they were asked (directly) without being presented with a rule before. Changes in the form of the rule by comparing "if *p* then *q*" with "*p* only if *q*" rules also affected inferences (Klauer et al., 2010). All these findings suggest that phrasing rules as either universal or existential should affect inferences.

In the following, we present two experiments on defeasible reasoning with quantifiers. In Experiment 1 we phrased conditionals from the literature as quantified rules and tested the participants' consideration of exceptions in a classical inference task. In Experiment 2 we use legal rules to investigate whether quantifiers also influence the inferences with emotionally charged content.

Experiment 1

Methods

Participants 41 participants took part in the experiment. One of them had to be excluded because he or she stated after the experiment to have had prior knowledge on formal logic. The remaining 40 participants were on average 21.50 years old (*SD* = 2.82).

Materials and Design For Experiment 1 we took 12 conditionals from the existing literature and phrased them with universal or existential quantifiers. 8 of the 12 conditionals came from De Neys, Schaeken and d'Ydewalle (2002), and 4 from Verschueren, Schaeken and d'Ydewalle (2005). According to the authors half of the conditionals have many exceptions, the other half few exceptions[1]. We rephrased these 12 conditionals either as statements with universal or existential quantifiers by adding either an "All" or a "There is at least one *Z* that" (replacing *Z* by the object of the rule) in the beginning of each statement. For an illustration see Table 1.

[1] In the literature authors make the distinction between disablers and alternatives. The former refers to situations that hinder the causal relationship between *p* and *q*. The latter refer to alternative situations which also bring about *q*, without the necessity of *p*. We selected our items only on the basis of disablers, because the amount of alternatives does usually not influence MP and MT inferences (cf. Cummins, 1995).

Table 1: Rephrasing conditionals with many and few exceptions to universal or existential rules in Experiment 1.

	Exceptions	
Rule	Many	Few
Conditional	If a person studies hard, then this person will do well on the test.	If a person jumps into the pool, then the person gets wet.
Universal	All persons that study hard will do well on the test.	All persons that jump into the pool get wet.
Existential	There is at least one person that studies hard and does well in the test.	There is at least one person that jumps into the pool and gets wet.

Each quantified rule was presented twice, once as a MP inference:

Rule (universal/ existential):
All persons that jump into the pool will get wet. /
There is at least one person that jumps into the pool and gets wet.
Fact:
Person X jumps into the pool.
Conclusion:
Person X gets wet.

And once as a MT inference:

Rule (universal/ existential):
All persons that jump into the pool will get wet. /
There is at least one person that jumps into the pool and gets wet.
Fact:
Person X does not get wet.
Conclusion:
Person X did not jump into the pool.

The person or object described in the fact was always labeled "X" (e.g., Person X, girl X, apple X) to emphasize that we are referring to one particular person or object. Participants had to indicate how strongly they accept the conclusion. The kind of quantifier was varied between individuals: 19 participants were confronted with universal quantifiers and 21 with existential quantifiers. Experiment 1 thus followed a 2 (exception: many vs. few) x 2 (inference: MP vs. MT) x 2 (quantifier: universal vs. existential) mixed design. The amount of exceptions and the kind of inference was varied within individuals. In total, participants solved 24 problems.

In addition to the inference task, we also included a generation task. Similar to De Neys et al. (2002; 2003a) and Cummins et al. (1991) participants had 1.5 minutes to generate exceptions for the different rules we used in the inference task ("A person jumps into the pool but does not get wet." [Why?]). This served to corroborate that our German translation of the rules did not alter the amount of exceptions participants can generate for each rule.

Procedure The experiment was programmed with Superlab 4.5 from Cedrus Cooperation. Participants were tested individually. In the instructions participants were told that they will be presented with statements containing some general rule and that their task is to indicate how strongly they accept a certain statement given the previous rule. Participants gave their answers on a 5-point-Likert scale ranging from no acceptance to full acceptance (the order of the extremes was counterbalanced). Each statement (the quantified rule, the fact, and the conclusion) was presented on a separate screen. Participants could switch to the next screen by pressing the space bar. The conclusion was written in red font and was followed – on a separate screen – by a figure of the 5-point-Likert scale where participants had to indicate their acceptance of the conclusion. We measured acceptance ratings and decision times. Participants were told to answer intuitively and that right or wrong answers do not exist (cf. Cummins, 1995; De Neys et al. 2003b). The 24 inference problems were presented in a random order after a short practice trial. After the inference task participants completed the generation task.

Results

Generation Task (Manipulation Check) Participants generated more exceptions for rules classified as having many exceptions (M = 4.9, SD = 1.3) than for those classified as having few exceptions (M = 3.3, SD = 1.0), $t(39) = 13.45$, $p < .001$, $d = 1.36$. In addition, the time needed to generate the first exception correlated significantly with the amount of exceptions generated, $r(38) = -.52$, $p \leq .01$.

Inference Task We conducted two separate 2 (exception: many vs. few) x 2 (inference: MP vs. MT) x 2 (quantifier: universal vs. existential) mixed Analyses of Variance (ANOVAs), once for acceptance ratings and one for decision times. Descriptive statistics are in Table 2.

For the ANOVA for acceptance ratings we coded full acceptance with 5 points and no acceptance with 1 point. We found a main effect of amount of exceptions, $F(1, 38) = 76.06$, $p < .001$, $\eta_p^2 = .67$, and a main effect of inference, $F(1, 38) = 49.95$, $p < .001$, $\eta_p^2 = .57$. Acceptance ratings were higher for rules with few exceptions (M = 4.2, SD = 0.5) compared to those with many exceptions (M = 3.3, SD = 0.7). And acceptance ratings were higher for MP (M = 4.1, SD = 0.5) than for MT inferences (M = 3.5, SD = 0.7). All other effects were not significant ($F \leq 1.52$, $p \geq .225$).

The ANOVA for decision times showed a main effect of inference, $F(1, 38) = 16.63$, $p < .001$, $\eta_p^2 = .30$, and a main effect of quantifier, $F(1, 38) = 4.77$, $p = .035$, $\eta_p^2 = .11$. Participants needed more time to decide for MT inferences (M = 2.2s, SD = 1.54) compared to MP inferences (M = 1.83s, SD = 1.36). They also needed more time to make a decision when the initial rule contained an existential quantifier (M = 2.45s, SD = 1.76) compared to a universal quantifier (M = 1.51s, SD = 0.60). All other effects were not significant ($F \leq 3.73$, $p \geq .061$).

Table 2: Acceptance ratings (AR) and decision times in seconds (DT) for universal and existential rules with many and few exceptions for Modus Ponens (MP) and Modus Tollens (MT) inferences in Experiment 1. Standard deviations are shown in brackets.

Rule	Many exceptions		Few exceptions	
	MP	MT	MP	MT
Universal				
AR	3.8 (0.8)	3.1 (1.1)	4.6 (0.4)	4.1 (0.8)
DT	1.6 (0.8)	1.7 (0.7)	1.2 (0.8)	1.5 (0.6)
Existential				
AR	3.5 (0.7)	3.0 (0.7)	4.4 (0.4)	3.8 (0.7)
DT	2.2 (1.5)	2.7 (2.0)	2.2 (2.6)	2.7 (2.4)

Discussion

Our results show that defeasible reasoning also exists when reasoning with quantifiers: participants accepted conclusions following from quantified rules with many exceptions less strongly than from quantified rules suggesting few exceptions. In addition, in accordance with the existing literature (Evans, 2002), participants accepted conclusions for MP inferences more often than for MT inferences. However, the kind of quantifier used in the initial rule did not affect acceptance ratings, but decision times: participants needed more time to select conclusions for existential than for universal quantifiers. One explanation for the similar acceptance ratings but different decision times is that existential quantifiers do not trigger a higher consideration of exceptions. Instead, the higher decision times for existential quantifiers might only reflect some kind of "translation" process. The wording "There is at least one person that…" might sound awkward in everyday language and thus participants probably needed extra time to "translate" the phrase into a more common wording. This translation might have resulted in the same inference pattern found for universal quantifiers. Another explanation is that, in a first step, the existential quantifier actually triggered a higher consideration of exceptions. However, since the existential quantifiers are known to create confusions and are difficult to understand (see Newstead, 1989), participants might have finally decided to ignore the quantifier and to answer in the way they usually do, e.g. calculating the conditional probability of q given p. To clarify this mismatch between acceptance ratings and decision times, in Experiment 2 we phrased the existential rules more naturally.

Experiment 2

In Experiment 2 we investigated the role of quantifiers in defeasible reasoning with legal rules. The legal rules could consist of either universal or existential quantifiers. We selected legal rules because such rules typically have exceptions. In law such exceptions are known as exculpatory circumstances and are reasons for voiding punishment, such as e.g., self-defense. Yet, recent studies show that laypeople usually ignore such exculpatory circumstances when an offence is highly morally outraging (e.g., Gazzo Castaneda

& Knauff, 2013). But what happens when the legal rules already suggest the possibility of exceptions? If the quantifiers used in the initial rule are considered during reasoning, then legal rules with existential quantifiers like "Some persons who kill another human being should be punished" should bring participants to apply the initial legal rule less often (and consider thus more exceptions) compared to rules with universal quantifiers like "All persons who kill another human should be punished".

Methods

Participants 43 participants took part in the experiment. Two participants had to be excluded because of technical problems and another one because the participant afterwards reported to study law. The remaining 40 participants were on average 23.08 years old (*SD* = 3.24).

Materials and Design For Experiment 2 we selected from a pilot study (*N* = 87 and *N* = 82) 6 high and 6 low morally outraging offences. These offences were embedded into quantified legal rules by either adding an "All" or a "Some" in the beginning of each statement (for an illustration see Table 3). We changed the phrasing of the existential quantifier compared to Experiment 1 to make the legal rules sound more naturally and thus avoid confounds with comprehension problems.

Table 3: Rephrasing high and low morally outraging legal rules as universal or existential rules in Experiment 2.

Rule	Moral Outrage of Offence	
	High	Low
Legal Rule	Whoever kills a human being, without being a murderer, is punished for manslaughter with imprisonment for not less than five years	Whoever organizes without governmental permission a game of chance or provides the facilities for this, is punished with imprisonment up to two years or with fine.
Universal	All persons that kill another human should be punished for manslaughter.	All persons that organize a game of chance without governmental permission should be punished for unauthorized organization of games of chance.
Existential	Some people that kill another human should be punished for manslaughter.	Some persons that organize a game of chance without governmental permission should be punished for unauthorized organization of games of chance.

As in Experiment 1, the kind of quantifier was varied between individuals (*n* = 19 got the rules with universal, and *n* = 21 with existential quantifiers) and each quantified rule was presented twice, once as a MP:

Rule (universal/ existential):
All persons that kill another human should be punished for manslaughter/
Some persons that kill another human should be punished for manslaughter.
Fact:
A person kills another human.
Conclusion:
Should the person be punished for manslaughter?

And once as a MT inference:

Rule (universal/ existential):
All persons that kill another humans should be punished for manslaughter/
Some persons that kill another human should be punished for manslaughter.
Fact:
A person is not punished for manslaughter.
Conclusion:
Did this person kill another human?

Note that contrary to Experiment 1 we phrased the conclusion as a question. After each inference, participants were told to rate their certainty on a 3-point-Likert scale (uncertain – neutral – certain). In total, participants were confronted with 24 problems. Experiment 2 followed thus a 2 (moral outrage: high vs. low) x 2 (inference: MP vs. MT) x 2 (quantifier: universal vs. existential) mixed design.

Procedure The experiment was programmed in Superlab 4.5 from Cedrus Cooperation. Participants were tested individually. In the instructions participants were told that we will confront them with statements describing legal cases, including some general rule about the offence and that they have to decide for each case whether they would apply the initial legal rule or not. Each statement (the quantified legal rule, fact, and conclusion) was presented on a separate screen. Participants could switch to the next screen by pressing the space bar. They gave their answer about the conclusion – which was written in red font – by either pressing a "Y" (yes) or "N" (no) key on the keyboard. The certainty ratings were given by pressing one of three keys from the numerical pad. We measured participant's conclusions, the decision times, and the certainty ratings. Participants were told to answer intuitively and that right or wrong answers do not exist. The 24 inference problems were presented in a random order after a short practice trial.

Results

We conducted three separate 2 (moral outrage: high vs. low) x 2 (inference: MP vs. MT) x 2 (quantifier: universal vs. existential) mixed ANOVAs, one for conclusions, one for decision times (corrected for sentence length) and one for certainty ratings. Descriptive statistics are in Table 4.

For the ANOVA for the conclusions we computed the percentage of logically "correct" responses per category (i.e., “yes” for MP and “no” for MT inferences). We found a main effect of moral outrage, $F(1, 38) = 28.37$, $p < .001$, $\eta_p^2 = .43$, but also an interaction between moral outrage and inference, $F(1, 38) = 34.87$, $p < .001$, $\eta_p^2 = .48$. For MP inferences, when the offence was of high moral outrage, participants selected significantly more often to apply the initial rule and to conclude that the offender should be punished compared to when the offence was of low moral outrage, $t(39) = 7.03$, $p < .001$, $d = 1.47$. In contrast, for MT inferences moral outrage did not affect the participants’ conclusions, $t(39) = 0.80$, $p = .430$, $d = 0.08$. All other effects were not significant ($F \leq 2.75$, $p \geq .106$).

The ANOVA for decision times showed a main effect of inference, $F(1, 38) = 25.82$, $p < .001$, $\eta_p^2 = .41$. Participants needed more time to decide for MT inferences ($M = 3.95$s; $SD = 1.56$) than for MP inferences ($M = 2.80$s; $SD = 1.08$). We also found an interaction between inference and moral outrage, $F(1, 38) = 4.68$, $p = .037$, $\eta_p^2 = .11$. For MP inferences participants were faster when the offence was of high moral outrage, but for MT it was the other way around. However, both post hoc t-tests did not reach the Bonferroni adjusted alpha level of .025 (ts ≤ 2.22, $p \geq .033$). All other effects were not significant ($F \leq 3.31$, $p \geq .077$).

The ANOVA for certainty ratings revealed a main effect of moral outrage, $F(1, 38) = 8.48$, $p = .006$, $\eta_p^2 = .18$, a main effect of inference, $F(1, 38) = 47.92$, $p < .001$, $\eta_p^2 = .56$, and an interaction between moral outrage and offence, $F(1, 38) = 35.25$, $p < .001$, $\eta_p^2 = .48$. For MP inferences participants were more certain when the rule contained a high moral outrage offence compared to when it contained a low moral outrage offence, $t(39) = 6.08$, $p < .001$, $d = 1.34$. For MT inferences, however, differences in certainty ratings for offences with high and low moral outrage did not reach the Bonferroni adjusted alpha level of .025, $t(39) = –2.14$, $p = .039$, $d = 0.21$. All other effects were not significant ($F \leq 2.01$, $p \geq .164$).

Table 4: Logically correct responses in percent (C), decision times in seconds (DT) and certainty ratings (CR) for high and low morally outraging universal and existential rules for Modus Ponens (MP) and Modus Tollens (MT) inferences in Experiment 2. Standard deviations are shown in brackets.

Rule	High Moral Outrage		Low Moral Outrage	
	MP	MT	MP	MT
Universal				
C	92.1(11.6)	84.2 (26.3)	60.5 (25.6)	89.5 (17.8)
DT	2.9 (1.1)	4.4 (1.6)	3.2 (0.8)	4.3 (2.3)
CR	2.9 (0.2)	2.1 (0.5)	2.4 (0.3)	2.2 (0.3)
Existential				
C	96.0 (9.0)	78.6 (34.6)	67.2 (26.3)	78.6 (33.8)
DT	2.2 (1.1)	3.8 (1.4)	2.8 (1.6)	3.4 (1.7)
CR	2.9 (0.2)	2.0 (0.7)	2.6 (0.4)	2.1 (0.6)

Discussion

Our results show that defeasible reasoning also exists when reasoning with quantified legal rules. For MP inferences with low morally outrageous offences participants defeated more often the logically valid conclusion to punish the offender than when the offence was of high moral outrage (and were less certain). MT inferences were not influenced by moral outrage. Moral outrage probably only affected MP but not MT inferences because only the conclusion of the former may activate personal values of the reasoner by asking what should happen to an offender. Contrary, the structure of MT inferences is less emotionally charged, because participants are only asked whether an offence was committed or not.

However, despite these results, we found no effects of quantifiers; not even in decision times. This lets us assume that the differences in decision times we found in Experiment 1 only resulted from the way in which we phrased the existential quantifiers and that the actual quantifier does not influence inferences at all.

General Discussion

Our results show that defeasible reasoning also exists with quantified statements. When a quantified rule suggests many exceptions people accept conclusions which are valid according to classical logic less often than when those rules suggest only few exceptions (Experiment 1). At the same time, the defeasibility of conclusions also depends on how emotionally attached one is to the initial rule (Experiment 2). However, contrary to our expectations, the consideration of exceptions was not moderated by the quantifier used in the rule. In fact, the kind of quantifier used in our tasks did not affect inferences. It seems that participants ignored the exact wording of the rule, but only extracted the topic of the task and used prior knowledge to indicate how highly they accept a certain conclusion. This interpretation would fit with the idea of what is known in the literature as System 1: a non-analytic and fast way of reasoning (e.g., Kahneman, 2011). However, such an explanation – as well as our results – conflicts with Klauers et al. (2010) dual source model, because the dual source model argues that the logical form influences inferences.

One explanation for the mismatch between our results and the ones observed by Klauer et al. (2010) is that the dual source model perhaps is simply not appropriate for quantified rules. Maybe, people do not represent the logical differences between universal and existential quantifiers mentally or at least not in a way it could affect inferences. Already Singmann et al. (2014) said that the component “logical form” of the dual source model does not refer to the actual logical status but to the “belief in the logicality of logical forms” (Singmann et al., p. 4). However, this explanation is not plausible because in our everyday lives we draw a distinction between all and some – even though this distinction does not perfectly correspond to the logical meanings of the quantifiers (Newstead, 1989).

Another explanation could be that the response formats we used in our experiments differ from the ones used in Klauer et al. (2010). In Klauer et al. participants had to rate how probable a certain conclusion is. In contrast, we asked our participants in Experiment 1 to indicate how highly they accept the conclusion, and in Experiment 2 to answer the conclusion by either selecting yes or no. Already Markovits, Forgues and Brunet (2010) showed that the response modality affects inferences. Yet, the fact that we found the same results in Experiment 1 and 2 – both using different response modalities – indicates that our findings are independent from response modality.

A third explanation for the mismatch between our findings and the ones from Klauer et al. (2010) is that maybe the form component postulated by the dual source model is less strong for quantified rules than for conditional rules. It is possible that the effect of quantifiers is too weak to be measurable with acceptance ratings. Further studies with mathematical models may help clarifying this question.

Investigating defeasible reasoning with quantifiers is novel and requires further investigation. For instance, in this study we only used the logically valid inferences MP and MT. Further studies could test the role of quantifiers in reasoning with inferences like the *Acceptance of the Consequent* and *Denial of the Antecedent* – which are often classified as valid by laypeople although being invalid according to classical logic. One could also manipulate the kind of quantifier within subjects. We decided to work with a between subjects design to avoid artefacts because of demand characteristics. However, if demand characteristics indeed evoke differences between universal and existential quantifiers, this would show that people know the differences between both quantifiers but that they do not care about these differences spontaneously.

We are aware that there are still open questions. However, we think that varying the wording and logical meaning of rules in inference tasks is promising: as we show in our experiments, it allows for testing the relative impact of background knowledge and logics in human reasoning.

Acknowledgments

This work was supported by DFG grants KN 465/10-1 and KN 465/10-2 to Markus Knauff within the DFG Priority Program "New Frameworks of Rationality". We thank Jessica Ewerhardy and Katharina Horn for data collection.

References

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology 38*, 191-258.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23,* 646-658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. Memory and Cognition, 19, 274-282.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition, 30,* 908-920.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory and Cognition, 31*, 581-595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology, 15,* 161-176.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128,* 978-996.

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning, 18*, 5-31.

Gazzo Castaneda, L. E., & Knauff, M. (2013, March). The effect of moral outrage and perspective on conditional legal reasoning. In U. Ansorge, E. Kirchler, C. Lamm, & H. Leder (Eds.), *Abstracts of the 55th Conference of Experimental Psychologists* (p. 93). Lengerich: Pabst.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A Theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 636-678.

Kahneman, D. (2011). *Thinking, fast and slow.* London: Allen Lane.

Klauer, K. C., Beller, S., & Hütter, M. (2010). Conditional reasoning in context: A dual-source model of probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 298-323.

Markovits, H., Forgues, H. L., & Brunet, M.-L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition, 38,* 485-492.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language, 28,* 78-91.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *TRENDS in Cognitive Sciences, 5*, 349-357.

Oaksford, M., & Chater, N. (2013) Dynamic inference and everyday conditional reasoning in the new paradigm, *Thinking & Reasoning, 19,* 346-379.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 26,* 883-899.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68,* B93-B101.

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology, 5:316*. doi: 10.3389/fpsyg.2014.00316

Verschueren, N, Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking and Reasoning, 11*, 239-278.

Weidenfeld, A., Oberauer, K., & Hörnig, R. (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *The Quarterly Journal of Experimental Psychology, 58A*, 1479-1513.